BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C12Q 1/68, D01F 4/02, C07K 1/34, 14/435

A1

(11) International Publication Number: WO 97/08315

(43) International Publication Date: 6 March 1997 (06.03.97)

(21) International Application Number: PCT/US96/13767

(22) International Filing Date: 22 August 1996 (22.08.96)

(30) Priority Data:
08/517,694 22 August 1995 (22.08.95) US

(71)(72) Applicants and Inventors: BASEL, Richard, M. [US/US]; 10760 W.C.R. 18, Fostoria, OH 44830 (US). ELION, Glenn, R. [US/US]; 442 Main Street, Chatham, MA 02633 (US).

(74) Agent: KALLAS, Nicholas, N.; Fitzpatrick, Cella, Harper & Scinto, 277 Park Avenue, New York, NY 10172 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CLONING METHODS FOR HIGH STRENGTH SPIDER SILK PROTEINS

(57) Abstract

This invention relates to methods of producing DNA fragments encoding silk proteins from silk-producing spiders. The present invention also relates to the DNA sequences encoding the spider silk proteins. This invention still further relates to methods of producing spider silk proteins using the above-described DNA sequences. The methods of cloning and producing proteins of the present invention are applicable to all silk-producing spiders. Clones developed by these methods produce commercially useful quantities of high molecular weight silk proteins. Because the silk made from such proteins have superior strength properties, the cloned silk proteins of the present invention are of considerable industrial importance.

```
                9                  18                  27                  36                  45                  54
5' ACA GGA AAC AGC TAT GAC CAT GAT TAC GAA TTC GGA TCC ATG GCA GCA GCA GCA
                                                          Met Ala Ala Ala Ala

               63                  72                  81                  90                  99                 108
GCA GCA GCT GGA GGT GCC GGA CAA GGA GGA TAT GGA GGT CTT GGA AGC CAG GGT
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly

              117                 126                 135                 144                 153                 162
GCT GGA CGA GGT GGA CAA GGT GCA GGC GCA GCC GCA GCA GCA GCC GGA GGT GCT
Ala Gly Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala

              171                 180                 189                 198                 207                 216
GGA CAA GGA GGA TAC GGA GGT CTT GGA AGC CAA GGT GCT GGA CGA GGA GGA TTA
Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Leu

              225                 234                 243                 252                 261                 270
GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA GCA GCA GCT GGA GGT GCC GGA CAA
Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln

              279                 288                 297                 306                 315                 324
GGA GGA CTA GGT GGA CAA GGT GCT GGA CAA GGA GCT GGA GCA GCC GCT GCA GCA
Gly Gly Leu Gly Gly Gln Gly Ala Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala

              333                 342                 351                 360                 369                 378
GCT GGT GGT GCC GGA CAA GGA GGA TAT GGA GGT CTT GGA AGC CAA GGT GCT GGA
Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly

              387                 396                 405                 414                 423                 432
CGA GGT GGA CAA GGT GCA GGC GCA GCC GCA GCA GCA GCC GGA GGT GCT GGA CAA
Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln

              441                 450                 459                 468                 477                 486
GGA GGA TAC GGT GGA CAA GGT GCC GGA CAA GGA GGC TAT GGA GGA CTT GGA AGT
Gly Gly Tyr Gly Gly Gln Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser

              495                 504                 513                 522                 531                 540
CAA GGT GCT GGA CGA GGA GGA TTA GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA
Gln Gly Ala Gly Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala

              549                 558                 567                 576                 585                 594
GCA GCA GCT GGA GGT GCC GGA CAG GGA GGA TTA GGT GGA CAA GGT GCT GGA CAA
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Leu Gly Gly Gln Gly Ala Gly Gln
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

TITLE

Cloning Methods For High Strength Spider Silk Proteins

Field of the Invention

This invention relates to novel methods of producing DNA fragments encoding for spider silk proteins. The present invention also relates to the DNA sequences encoding the spider silk proteins. This invention still further relates to novel methods of producing spider silk proteins using the above-described DNA sequences. The invention also relates to methods of purifying these spider silk proteins and manufacturing fibers and films from them.

Clones developed by the methods of the present invention produce commercially useful quantities of high molecular weight spider silk proteins ranging in molecular weights from 90,000 to over 250,000, which are from 40% to greater than 100% of the molecular weight of natural major ampulate (dragline) spider silk protein obtained from Nephila clavipes. Because the silk made from these high molecular weight proteins have superior physical properties, such as high tensile strength and substantial elasticity, the cloned silk proteins of the present invention are of considerable industrial importance.

These spider silk proteins have been cloned by several methods of the present invention and the natural sequence spider silk clones have been produced in E. coli expression systems. These expression systems have then been used to produce various partial and full length natural spider silk proteins, which have been expressed at levels in excess of 2 grams per liter of cell mass. These spider silk proteins are then purified and used for many purposes such as spinning fibers, forming films and other applications resulting from the weaving of filaments.

Background of the Invention

Silk production by many diverse animal orders (e.g., insects, arachnids and mites) is well known. Spiders, for example, produce natural webs and draglines having high tensile strengths. Silkworms, on the other hand, although producing silks at high production rates, have silk proteins that are considered inferior to spider silk proteins in their physical properties. For example, silkworm proteins have considerably lower tensile strengths than spider silk proteins. Orb weavers and other spiders, although naturally producing low quantities of silk filaments (less than economic for commercialization), have strong filaments. In fact, spider filaments can be several times stronger than Kevlar® (9.5×10^4 vs 3×10^4 Jkg^{-1}). These superior strength properties make spider silk protein filaments a preferred choice for parachutes, sails, body armor and other high strength applications requiring strong filaments. Additionally, these spider filaments find utility as absorbent films for many heavy metals and organics including biological weapons. They also find utility as absorbents that selectively bind DNA and absorbents for many other chemicals, flavors and fragrances.

Although it might be hypothesized that spider silk could be produced from culturing spiders, this is impractical for several reasons. First, in addition to being very difficult to raise, spiders will eat their neighbors if grown in very high densities. Second, spiders produce only small amounts of silk protein making production of even milligram quantities prohibitively expensive. As a result of these limitations, the only acceptable method for producing commercial quantities of spider silk proteins is to clone the spider gene into an acceptable large scale production vector. The present invention accomplishes that objective.

Synthetic silk protein genes have previously been produced by making short base pair segments and then using large numbers of repeating units. Proteins with modest molecular weights (ranging from 20,000 to 80,000) have been obtained by such a process. To achieve a variety of physical properties, this process has been varied and synthetic proteins with different sequences have been produced. For example, prior workers have used sequences obtained by taking small lengths of naturally occurring silk proteins and changing the sequences.

These prior methods, however, have resulted in materials that are inferior to natural silks. Moreover, in some cases, this prior technology has also produced clones that are unstable for the long term use required for commercial applications. Therefore, one of the objects of the present invention is to overcome the above-mentioned problems that occur with polymerized short DNA sequences. This is accomplished with the present invention by the production of long DNA that encode for high strength, high molecular weight silk proteins.

Because of the potential that high strength major ampulate (dragline) spider silk offers, silks from orb weavers such as Nephila clavipes have been studied in attempts to understand the molecular basis of their strength. Researchers have also attempted with limited success to clone the natural protein or make a synthetic silk gene by incorporating the repetitive elements responsible for the high strength of spider silk fibers.

There are, however, many problems associated with cloning a silk protein. First, the natural protein amino acid sequence is composed of numerous repeating subunits, and therefore does not have many unique sites that can be used to clone the natural gene. The literature indicates that the carboxy end of dragline spider silk protein from Nephila clavipes is the only area shown to be unique. This has lead to only a few prior attempts at cloning the natural gene, and consequently many more prior attempts at making a synthetic protein. Nevertheless, making stable clones and the resultant synthetic spider silk proteins are replete with problems owing to the repetitiveness of the DNA sequence that is being mimicked. For example, DNA with high amounts of repeats (especially GCA repeats) is unstable due to transcription errors and the high probability of recombinational deletions, resulting in constantly changing DNA. Because of these problems, the integrity of many clones has been questionable.

In nature, silk genes are quite stable because of the intermixing of repetitive and non-repetitive regions. Unfortunately, synthetic genes do not lend themselves to such constructions as they are highly unstable to recombination and recombinational deletions in particular. Failure to obtain stable clones has

occurred because insufficient amounts of the gene were cloned and therefore the non-repetitive regions that possibly form the basis of their natural stability were not obtained. It is therefore one of the objects of the present invention to obtain stable clones having non-repetitive as well as repetitive regions.

Another problem with cloning dragline spider silk is the size of the gene. Spider silk proteins typically are 200,000 kDa or higher and the corresponding genes also have at least one intron. As such, it is projected that the size of the DNA fragment would be in the range of 5-10 Kb plus any introns. With current technology, genes of this magnitude are still notoriously difficult to clone. The present invention has overcome this problem.

Because of their mechanical strength properties, much attention has been directed to the cloning of spider silk proteins. A major advance in understanding Nephila clavipes dragline silk was taken by Xu et al. (Proc. Natl. Acad. Sci. 87:7120, 1990). Xu et al. ascertained a portion of the repetitive sequence of a spider dragline silk from a partial clone. Although this repeating unit encoded for up to 34 amino acids, it was not exactly conserved as the sequence had deletions and changes in some of the repeats. Nevertheless, Xu et al. discovered two important areas in the sequence -- repetitive regions which give spider silk some of their properties and a non-repetitive (carboxy) region. Hinman and Lewis (J. Biol. Chem. 267:19320, 1992) reported a second cDNA clone presumed to be from a second spider protein. This sequence had a similar repetitive region as that discovered by Xu et al. and a carboxy terminal non-repetitive end. The Hinman and Lewis repeating unit was longer, encoding for 51 amino acids, and highly variable.

In the expression of spider silk proteins by Lewis et al., European Patent Application EP 0452925 A2, published 10/23/91, only small protein fragments were apparently produced in small yields. These small protein fragments are probably of no commercial value because good mechanical properties only result from larger proteins, especially those close to full length. Lombardi et al., International Patent Application WO 91/16351, published 10/31/91, also produced a recombinant spider silk protein in very low yields, but these clones appeared to have low mechanical strength due to their small molecular weights.

It is also theorized that the spider silk clones heretofore developed do not represent faithful copies of the natural gene. This is confirmed by a number of studies, for example, Beckwith & Arcidiacono (J. Biol. Chem. 269(9):6661, 1994) showing that both spider proteins have a high homology and may in fact represent the same protein.

Although many researchers have conceded that natural expression systems are useful as silk variants, they have been unable to overcome the expression problems based upon codon preferences. While it is believed that using highly conserved repetitious repeat regions can produce improved proteins, these synthetic gene expression systems suffer from DNA stability problems, low expression rates, and the production of proteins with less desirable properties than those of natural spider silk. It is therefore an object of the present invention to overcome these problems.

Ferrari et al., International Patent Application WO 88/03533, published 5/19/88, disclosed synthetic genes which produced protein with silk-like properties. In addition, a number of small repeat proteins mimicking

natural fiber proteins were developed by Cappello et al., International Patent Application WO 90/05177, published 5/17/90. Floyd, International Patent Application WO 94/29450, published 12/22/94, also attempted to develop a spider silk synthetic gene using a number of natural repeat units developed by Xu et al. All of these clones, however, have small molecular weights that preclude them from having the desired properties of natural spider silk.

The present invention relates to the novel synthesis of partial and full length spider silk protein clones. Some of these partial length clones have also been multimerized into other clones with molecular weights up to and exceeding those of natural spider silk. The present invention has made it possible to develop natural silk-like clones that have a complete range of properties. One skilled in molecular biology can use these clones as a starting point for creating clones with other useful silk properties such as strength, yield point, adhesiveness and plasticity. Furthermore, these new sequences can be used as starting points to design other synthetic genes. For some spiders which incorporate colors or pigments into their silk proteins, these methods may also permit naturally colored protein.

The present invention also relates to unique chemical methods for fermentation of transfected hosts in culture media. One of the major problems of producing silk proteins by bacterial fermentations is the partial digestion of proteins by proteases. In fact, the rate of protein decomposition from proteases can in some cases overcome the rate of high molecular weight silk protein expression, thereby making commercial operations impractical. The present invention overcomes this potential problem.

These and additional objects and advantages of the present invention are shown from the descriptions below.

Brief Description of the Figure

Figure 1 shows the 2Kb DNA sequence for encoding the spider silk protein.

Summary of the Invention

This invention relates to a process of producing DNA fragments encoding for silk protein, comprising the steps of (i) selecting target DNA harvested from a silk-producing spider, the target DNA comprising a plurality of repetitive and non-repetitive regions; (ii) selecting a single strand DNA primer of at least 10 nucleotides having a DNA sequence that is complementary to a region in the target DNA; and (iii) repetitively combining the DNA primer with melted target DNA and incubating the combined DNA primer and target DNA with nucleotides and a DNA polymerase having proofreading ability to produce the DNA fragment, wherein the DNA fragment is complementary to said target DNA and is at least 2 Kb. In a more particular embodiment, DNA fragments of at least 5 Kb can be produced.

In a further embodiment of the above-described process of producing a DNA fragment encoding silk protein, the process comprises the step of using two different DNA primers instead of one. In still further embodiments of the processes for producing a DNA fragment using a single strand DNA primer or two different DNA primers, the target DNA is cDNA made by reverse transcription of full length mRNA coding for spider silk, and the process further comprises the steps of (i) adding a

primer site to the amino terminal end of the first strand cDNA made thereof and (ii) using the poly T region of the cDNA as a first polymerase priming region. In a still further embodiment of these processes for producing a DNA fragment, a second primer site is created at the unknown end of the DNA using a ligation cassette. In a still further embodiment, a second primer site is created at the unknown end of the DNA using a terminal transferase to make a primer site selected from the group consisting of poly dT, poly dA, poly dG and poly dC.

The DNA primer for the above-described processes of producing a DNA fragment can be selected from DNA represented by starting and ending sequences (i) - (xx) given below:

GGCGAATTCGGATCCATGGCAGCAGCAGCAGCAGCAGCT;

(ii) GGCGAATTCACCCTGGGCTTGATAAACTGATTGAC;

(iii) GCATGCACGCATGGTGCATGGATGC;

(iv) TTCGAATTCATGGGCCCTGGACAACAAGGACCATCTGGACCT;

(v) GGAAGGCGGGCAGTGAGCGCAACGCAATTAATG;

(vi) GAYGAYGGNAAYGCNGT;

(vii) TGNTGNCCSGTTCG;

(viii) CGSCGKCGSCCACGSCCSCG;

(ix) GTTAAATGTAAAATCAAGAGTTGCTAA;

(x) GGCCAATCTCTTTTGAGTGCATTTTAA;

(xi) TAAGCAACTCTTGATTTTACATTTAAC;

(xii) TTAAAATGCACTCAAAAGAGATTGGCC;

(xiii) TCAGCAGAATCTGGACAACAAGGCCCA;

(xiv) CCNCGNCCNCTYCC;

(xv) GGTGCAGCAGCAGCAGCTGCWGG;

(xvi) GGTGGTGCCGGACAAGGAGGMTATGGAGGWCTTGGA;

(xvii) GGWGGACGAGGTGGATTA;

(xviii) GATAAAAAGAAATATGCTGCAGAACTTCACTTGGTTCAC;

(xix) CARGCNGGNGCNGCNGSNGGNGGNTTYGGNCC; and

(xx) GGNGGNGGNGCNGGNCARGCNGGNGCNGCNGSNGGNGGNTTYG GNCCNGGNGCNGGNGGN,

wherein N = G, A, T, C; V = G, A, C; B = G, T, C; H = A, T, C; D = G, A, T; K = G, T; S = G, C; W = A, T; M = A, C; Y = C, T; and R = A, G.

In a still further embodiment of the processes for producing a DNA fragment, the target DNA is selected by hybridization to a DNA probe, having at least one of the above-described sequences (i) - (xx), that is reversibly bound to a support to enrich for the silk-encoding DNA fragments.

In another process embodiment of producing a DNA fragment encoding silk protein, called the multimerization process, the process comprises the steps of (i) selecting a target DNA encoding silk protein harvested from a silk-producing spider, the target DNA comprising a plurality of repetitive and non-repetitive regions; (ii) selecting a first pair of different DNA primers, the first pair of DNA primers both being complementary to a region in the target DNA, and at least one of the first pair of DNA primers being represented by the sequences (i) - (xxvi); (iii) producing a first DNA fragment by repetitively combining the first pair of DNA primers with melted target DNA and incubating the combined DNA primers and target DNA with nucleotides and a DNA polymerase having proofreading ability to produce the first DNA fragment, the first DNA fragment being complementary to the target DNA and at least 2 Kb. This multimerization process further comprises the steps of (iv) selecting a second pair of different DNA primers, at least one of the second pair of DNA primers being different than both of the sequences of the first pair of DNA primers, and at least one of the second pair of DNA primers being represented by the sequences (i) - (xxvi); (v)

producing a second DNA fragment by repetitively combining the second pair of DNA primers with melted target DNA and incubating the combined DNA primers and target DNA with nucleotides and a DNA polymerase having proofreading ability to produce the second DNA fragment, the second DNA fragment being different than the first DNA fragment and also being complementary to the target DNA, the second DNA fragment being at least 2 Kb; (vi) restricting the first and second DNA fragments; and (vii) recombining the restricted portions of the first and second DNA fragments into a multimerized DNA, the multimerized DNA encoding spider silk protein and being at least 4 Kb in length.

In a more particular embodiment of the above-described multimerization process, all DNA primers are represented by sequences (i) - (xxvi). In another particular multimerization process embodiment, all DNA primers are different. In a still more particular multimerization process embodiment, the multimerized DNA is at least 6 Kb or 8 Kb in length.

In a DNA sequence embodiment, this invention relates to a DNA sequence encoding spider silk protein, wherein the DNA sequence comprises a plurality of repetitive and non-repetitive regions and has a length of at least 2 Kb. In a more particular embodiment, the DNA sequence has a length of at least 5 Kb. In a still more particular DNA sequence embodiment of the present invention, the DNA comprises the sequence illustrated in Figure 1.

In a process of producing silk protein embodiment, this invention comprises the steps of (i) selecting a DNA; (ii) inserting the DNA into an expression vector; (iii) transfecting host cells with the expression vector; (iv) fermenting the transfected host in culture media

to produce silk protein; and (v) recovering the silk protein. In a more particular silk protein production process embodiment, the culture media for fermenting the transfected host contains protease inhibitor. In a still further silk protein production process embodiment, the process comprises the steps of (i) applying ultrasound energy to rupture the host cells; (ii) applying ultrasound energy to resuspend the silk protein; and (iii) centrifuging the ruptured host cells to separate cell membranes from the silk protein. In these silk protein production processes, purification of the silk protein is accomplished by ultrafiltration or alcohol precipitation.

In a process for spinning silk protein embodiment, this invention relates to a process comprising the steps of (i) concentrating silk protein purified by ultrafiltration or alcohol precipitation; (ii) drawing a fiber of concentrated silk protein; (iii) spinning silk fibers to produce a silk thread; and (iv) washing the silk thread to remove any solubilization reagents. The solubilization reagents are selected from the group consisting of hexafluoroisopropanol, sodium hydroxide, potassium hydroxide, urea, urea phosphate, lithium salts, organic solvents, guanidine hydrochloride, ammonium sulfate, acetic acid, phosphoric acid, dichloroacetic acid, formic acid and sulfuric acid. In a still more particular process of spinning silk, the process further comprises the step of coating the silk fiber or thread with oxides of tin or titanium.

In a fabric embodiment, the present invention relates to a fabric comprising the spider silk threads made according to any of the processes of the invention. In a further fabric embodiment, the fabric comprises spider silk threads made in accordance with any of the processes of the present invention in combination with

Kevlar®, graphite or carbon fibers, as well as silkworm silk.

The protein can be used as a coating, extruded into a fiber, or made into a polymeric film.

Detailed Description of the Invention

Sources of Silk-Producing Spider DNA

While the methods of the present invention were specifically developed to clone Nephila clavipes major ampulate (dragline) spider silk, the methods of cloning and producing silk proteins are applicable to all silk-producing spiders.

As a group, spiders may have up to eight kinds of silk glands. Although no spider species has all eight silk glands, all spiders have at least three such glands and most have five. Each gland produces a different type of silk having different properties. For example, some silk dries quickly, while other silk remains sticky.

Spiders belong to the phylum Arthropoda, class Arachnida and order Araneae. True spiders belong to the suborder Labidognatha. Other spider types include comb footed, crab, fisher, funnel web, hackled-band, orb weavers, jumping and ogre faced stick. Spiders from any of the following genus groups can be used in accordance with the present invention: Micrathena, Mastophora, Metepeira, Araneus, Argiope, Nephila or Gasteracantha.

Orb weavers are among the most successful spider groups because they have evolved silks with remarkable strength and flexibility. The orb weavers are known as Argiopidae and include: arrowheaded shaped Micrathena

sagittata, bolas spider Mastophora cornigera, labyrinth Metepeira labyrinthea, marbled Araneus marmoreus, black-and-yellow garden Argiope bruennichi, golden silk Nephila clavipes, and spiny bodied Gasteracantha cancriformis.

Nephila clavipes has been studied the most in genetic research since its silk threads are strong and its silk glands are large and easy to dissect. Other orb weavers also produce strong silk threads.

While all spiders produce silk, the proteins that form the silk threads vary considerably in their molecular makeup and serve a variety of purposes. For example, the Antrodiatus spiders spin a simple kind of silk comprising just two proteins. In contrast, spiders in the family Araneoidea, called web spinners, produce up to eight different kinds of silk. Orb weavers produce a variety of silks using several proteins to create webs of greater strength and flexibility.

Spider silk proteins also have different qualities depending upon which silk gland it was spun from. The strongest silks known are from the major ampulate gland of orb weavers. Of the eight types of silk produced by orb spiders, the major ampulate (dragline) silk was selected for this work because of its physical strength and non-sticky properties. This dragline silk is composed of protein although carbohydrates are associated with the fiber. In the spider's spinneret, the liquid silk undergoes an irreversible transition to an insoluble form composed of a high relative ratio of alanine and glycine. This fiber consists of an antiparallel β-sheet with elastic interspaces. The amino acid composition of this silk (shown in the table below) mimics the composition of clones of the present invention.

Percent Amino Acid Composition Of
Nephila clavipes Major Ampulate Spider Silk

Amino Acid	Protein	220 kDa Band	190 kDa Band
glutamic acid	8.52	9.77	9.35
serine	3.51	2.57	2.79
glycine	41.66	45.88	44.80
arginine	1.28	1.98	2.28
alanine	25.25	28.57	28.35
proline	0.78	0.37	0.51
tyrosine	4.20	3.25	3.26
leucine	4.82	4.62	4.48

Silk Polymers, ACS, Symposium, Ser. 544, 1994.

Cloning

Two Primer PCR Cloning

Although many researches have tried cloning repetitive silk genes using PCR-type techniques, at least two problems have occurred. These PCR techniques could not transcribe DNA with good fidelity for a gene that was 8-15 Kb in length. In fact, most clones reported in the literature have been transcribed incorrectly. Therefore, the present inventors set out to overcome these shortcomings and found that by using somewhat degenerate primers either one or a number of PCR products could be produced.

Genomic DNA taken from Nephila clavipes abdomens was used. To isolate the DNA from the spider, the preparation method described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Vol.1-3, Cold Spring Harbor Laboratory, New York (1989), was followed exactly. This procedure resulted in high molecular weight genomic DNA in excess of 2 Kb.

The inventors experimented with many primers that were related to the sequence data disclosed by Xu et al. Some of the primers used are disclosed above as primer

sequences (i) - (xx). Although these primers were also tried by Beckwith & Arcidiacono, the present inventors are the first to produce spider silk protein up to 2 Kb in length using a two primer PCR cloning system. The present inventors were also able to produce spider silk proteins with higher Kbs by the claimed cDNA and single site cloning methods described below.

Initial conditions for PCR clones were produced using primers derived from spidroin 1 as defined by Xu et al. and Hinman and Lewis. Using normal PCR with Taq polymerase (Stratagene product no. 600131 under license from Perkin Elmer, Stratagene, 11011 North Torrey Pines Road, La Jolla, CA), the inventors could only get PCR products of up to 700-1000 bp, which supports the findings of others. Even these small pieces were considered of dubious quality. Using a Taq extender (Stratagene product no. 600148), a number of bands of up to about 1900 bp were obtained as shown in Example 1. However, when another polymerase with proofreading activity was used (Takara Taq LA), only one primary band was obtained as described in Example 2 below.

Example 1: Cloning with Taq polymerase

In this example, Nephila clavipes DNA isolated by the procedure of Beckwith & Arcidiacono was used along with the following two primers:
primers (i) GGCGAATTCGGATCCATGGCAGCAGCAGCAGCAGCAGCT,
and (ii) GGCGAATTCACCCTAGGGCTTGATAAACTGATTGAC.
Primer (i) codes for a poly-alanine repeat sequence based on the forward reading frame. Leader sequences that insert an in-frame start codon and both BamH I and EcoR I leader restriction sites for cloning as overhangs were also put into the primer. Primer (ii) is a PCR primer (bp 2218 to 2242) based upon the reverse sequence of Xu et al. This sequence also has

an in frame stop codon and an EcoR I restriction site. As shown in this Example and Example 2, the results depend on the PCR conditions and are not positive without newer polymerases. The regular Taq and the Taq extender did not give the same results, presumably due to misreading or false priming.

The PCR mix was as follows: 5 μl Taq extender buffer (Stratagene); 1 μl of Taq polymerase 5 μg/μl (Stratagene); 1 μl of 1μg/μl DNA template (spider genomic DNA); 1 μl of 2 μM primer (i) in water; 1 μl of 2 μM primer (ii) in water; 5 μl of NTP's (2 μM each of dATP, dCTP, dGTP, and dTTP, pH 7.0); 45 μl of Taq extender (Stratagene); and water to a total of 100 μl total.

The PCR cycler conditions were as follows: initial dwell 94°C. for 2 min; and PCR conditions (30 cycles): annealing at 60°C. for 1 min.; extension at 72°C. for 2 min.; and denaturation at 94°C. for 1 min. Alternatively, the PCR conditions of annealing at 60°C. for 1 min. and extension at 72°C. for 2 min. can be replaced with a treatment of 72°C. for 2 min.

A 5μl portion of this reaction mixture analyzed by 1% agarose electrophoresis showed DNA bands. With this technique, up to 7 DNA bands were achieved which were assumed to represent a number of alanine repeat regions in the sequence. The largest DNA fragment was 1900-2000 bp (and was referred to as a 2 Kb piece). This is essentially the same band as achieved in Example 2. This was cut out of the gel with Gene Capsule™ (Cat. No. 786-001 from Geno Technology Inc., 3830 Washington Blvd., St. Louis, MO 63108), purified with phenol and ethanol precipitation. These bands were cloned into E. coli XL1 MRF' super-competent cells using the procedure described in Example 2 below. The

2 Kb piece was also found when the Takara Ex Taq LA polymerase PCR conditions described in Example 2 were used with primers (ii) and (iii).

Example 2: Cloning with Takara Taq LA polymerase

The genomic DNA was isolated from freeze dried spider abdomens which were ground in a mortar and pestal and extracted according to Sambrook et al., Molecular Cloning: A Laboratory Manual Vol. 1-3, Cold Spring Harbor Laboratory, New York (1989).

The cloning for this Example was accomplished with the following primers:
primer (iii) GCATGCACGCATGGTGCATGGATGC, and
primer (ii) GGCGAATTCACCCTAGGGCTTGATAAACTGATTGAC.
Primer (iii) was made from the peptide sequence 4 described by Mello et al., Silk Polymers, ACS, Symposium, Ser 544 (1994). Primer (ii) was made as described in Example 1 above.

The following PCR mix and conditions were used. PCR mix: 5 μl 10X Takara LA PCR buffer; 5 μl Takara dNTP mix; 1 μl primer (iii) (2 μM); 1 μl primer (ii) (2 μM); 1 μl Takara Ex Taq with proofreading activity; 1 μl spider genomic DNA; water to a total of 50 μl; and 50 μl mineral oil. The Takara LA PCR buffer, dNTP mix, and Takara Ex Taq were supplied with a Takara Roll kit distributed by Panvera Corp., 565 Science Dr., Madison, WI 53711. PCR cycler conditions were as follows: initial dwell 94°C. for 1 min.; PCR conditions (30 cycles): annealing and extension at 68°C. for 1 min. and denaturation at 94°C. for 1 min.; and post dwell at 4°C.

To insert this 2 Kb piece into E. coli, a familiar vector, pUC18, was chosen because the plasmid had a good number of cloning sites and could express proteins well. It is also known that this vector is suited to sequence analysis using well known primers. To insert this 2 Kb piece into E. coli XL1 MRF', pUC18 was first prepared in a 1 μg/μl DNA preparation obtained from Sigma Chemical Co., P.O. Box 14508, St. Louis, MO 63178-9916. Restriction enzymes were also similarly used to digest the insert. The restriction protocol was as follows: 5 μg or less of plasmid or insert DNA; 5 μl of restriction enzyme 10X buffer; 5 μl 1 mg/ml acetylated BSA; 5 μl restriction enzyme (EcoR I); water to a final volume of 50 μl; and incubate for 3 hr. at 37°C.

The vector was also treated after phenol extraction and cleanup with EcoR I restriction enzyme. The vector was similarly treated with calf intestinal alkaline phosphatase (CIAP). This treatment prevented the vector from re-annealing.

The CIAP protocol, which was done in addition to the restriction protocol, was as follows: 10 μl CIAP 10X buffer consisting of 500 mM tris-HCl, pH 9.0, 10 mM $MgCl_2$, 1 mM $ZnCl_2$ and 10 mM spermidine; 1 unit CIAP; water to final volume of 100 μl; and incubate for 60 min. at 37°C. One CIAP unit will hydrolyze 6.0 mM of p-nitrophenyl phosphate per minute at 37°C. These units are measured in a 0.1 M glycine buffer at pH 10.4 containing 1.0 mM $ZnCl_2$, 1.0 mM $MgCl_2$. The next step was to ligate the insert into the pUC18. To do this, the DNA was repurified with phenol extraction and ethanol precipitation and then ligated according to the protocol described below.

Ligation protocol: 100 ng vector DNA; 100 ng or less insert DNA; 1 unit T4 DNA ligase (Weiss Units); 1 μl ligase 10X buffer; water to a final volume of 10 μl; and incubate for 1 hr. at room temperature.

The new vector was then inserted into E. coli XL1 MRF' obtained from Clonetech Laboratories, Inc., 4030 Fabian Way, Palo Alto, CA 94303, using the Clonetech method for inserting supercompetent cells. The transformants were selected by ampicillin resistance in LB broth 10 g/l bactopeptone, 5 g/l yeast extract, and 5 g/l NaCl using 50 μg/μl of ampicillin. Clones were checked for the proper insert by first looking for the proper size of plasmid, approximately 4.3 Kb. The insert was also checked by using biotinylated probes and assaying for hybridization. The best 5 inserts from transformation were checked for expression of the inserted protein as it was inserted in such a way that it should express within pUC18.

The 2 Kb insert was easily made using the PCR technique described above. This technique produced superior results over the following three methods: screening of shotgun clone libraries for silk by probes based upon peptide sequencing (Xu et al.); cDNA inserts from the silk gland (Hinman and Lewis); and PCR using Taq polymerase or other polymerases with no proof reading. (Beckwith and Arcidiacono).

The PCR technique of Example 2 compared to the above three methods was fast, did not induce errors into the sequence as was apparent from the other reported methods, and was directed only to the gene of interest. With just a little of the sequence from the amino end and carboxy end of the spider silk, this technique could be applied to the sequencing of silks other than

the major ampulate (dragline) silk or to other spiders having similar properties.

To determine whether the protein was expressing in the E. coli host, antibody assays were developed for the determination of spider silk protein. These antibody assays are discussed below. In addition, SDS gel electrophoresis, indicated that the 2 Kb insert was producing a 94 kDa protein in good yield. The gel electrophoresis was done according to the procedure of Mellow et al., Silk Polymers, ACS, Symposium Ser. 544 (1994). Using LB broth, the yields ranged from 0.1-10% of the total protein produced by the bacteria. Western blotting using BioRad Kit #170-6460 from Bio-Rad Laboratories, 3300 Regatta Blvd., Richmond, CA 94804, also confirmed that this protein was a silk protein, and it was the only protein showing antibody reaction.

This 2 Kb insert, as well as the other inserts developed by the present inventors, were sequenced according to normal conditions using the Promega silver sequence system, Cat. No. Q4130 Promega Corporation, 2800 Woods Hollow Rd., Madison, WI 53711-5399. This was done by using multiple primers, deletion clones and other clones based upon Example 1. This sequence has been characterized by the DNA and protein sequence shown in Figure 1.

Cloning from cDNA

Many researchers have attempted to clone Nephila clavipes spider silk with little success as only small pieces have ever been cloned. The problems associated with cloning from cDNA have included the inability to obtain full length mRNA, poor reverse transcription of the protein and poor fidelity. Another major problem has been the inability to obtain satisfactory amounts

of mDNA from the silk glands. The cDNA cloning technique of the present invention, which is described below, overcomes these problems.

Example 3

A. Development of full length mRNA

Before this cloning technique could be successful, the problem of obtaining full length mRNA had to be resolved. Since the copy number of mRNA in the silk gland of spiders is extremely low, it was decided to use the silkworm Bombyx mori in order to develop an analogous method of obtaining full length mRNA from spider silk glands. It was discovered after numerous mRNA isolation methods were tried, that a mRNA purification kit (# 8-MB4003K) from PerSeptive Diagnostics (Cambridge, MA) could consistently separate essentially full length mRNA without any appreciable degradation. This mRNA purification technique uses biomagnetic bead separation and oligo $(dT)_{20}$ particles to separate the mRNA.

B. Development of a long and accurate PCR technique

The next step in the development process was to convert the mRNA to a good first strand template and then reliably replicate the DNA. Using an Invitrogen Cycle mRNA reverse transcription, cDNA cycle kit L1310-01 obtained from Invitrogen Corp., 3985 B Sorrento Valley Blvd., San Diego, CA 92121, and a PCR amplification system proved unsatisfactory because the primers developed were only good for amplifying small pieces of mRNA. The inventors thereafter decided to develop their own technique for obtaining a 10 Kb mRNA. The first part of this process was to optimize the reverse transcriptase reaction. The preferred reverse

transcriptase for making the first strand was discovered by trying various reverse transcriptase enzymes, including AMV (Avian Myelobastosis Virus) reverse transcriptase (M5101) and M-MLV (Moloney Murine Leukemia Virus) reverse transcriptase (M5301) which is modified to remove the ribonuclease H activity. See Tanese & Goff, Proc. Natl. Acad. Sci. U.S.A. 85:1977 (1988). Both M5101 and M5301 were obtained from Promega Corp., 2800 Woods Hollow Road, Madison, WI 53711. The M-MLV used according to Promega instructions was preferred as it gave the highest fidelity and the longest product length. It is therefore recommended to use the M-MLV and the following reverse transcriptase protocol: 2 μl 10X reverse transcriptase buffer; 2 μl M-MLV reverse transcriptase (Promega); 2 μl dithiothreitol; 1 μl poly $d(T)_{20}$; and 13 μl mRNA.

After the first strand was created, it was necessary to amplify the mRNA piece (after it was reisolated by phenol extraction and ethanol precipitation). Since mRNA has a poly A end, a poly T primer was used. At the other end, a marker sequence was needed and numerous possibilities existed. While putting a marker cassette on each end worked, that technique had a low probability of ligating on to the low number first strand DNA. Since mRNA has a poly A end adjacent to where the carboxy end of the protein is coded, a method to label one end was already available. Therefore, a method that would just label the one end was adopted and a terminal transferase was used. The preferred method is to use the enzyme terminal transferase to add poly A at the 3' end of the first strand. This was done by allowing a single primer method to amplify both ends of the cDNA from the mRNA. The protocol is as follows: 10 μl terminal transferase buffer (Promega formula); 1 μl terminal transferase (Promega); 5 μl of

the first strand DNA from reverse transcription procedure described above; 1 μl oligo $d(T)_{6-12}$; 1 μl d(A); and 7 μl water; and incubate for 1 hr. at 37°C. Both the terminal transferase buffer and the terminal transferase were obtained from, Promega Corp., 2800 Woods Hollow Rd., Madison, WI 53711-5399, catalog no. M1871.

The DNA was then reisolated using phenol and ethanol precipitation, and PCR was used. The technique, which is described below, yielded DNA strands with a poly dA strand on one end and a poly dT on the opposite end. The problem of using PCR on such a long piece of DNA, which required long and accurate amplification protocol of the cDNA using poly T as the primer, was solved using the following Takara LA method of DNA amplification. The PCR amplification of cDNA was as follows: 1 μl DNA from the terminal transferase procedure described above; 10 μl 10X Takara LA buffer; 10 μl dNTPs (Takara); 1 μl poly $d(T)_{20}$ primer; 1 μl Takara Ex Taq LA polymerase; 78 μl water; and 100 μl mineral oil. The PCR conditions were as follows: the initial dwell was 94°C. for 1 min.; the amplification cycles (40) were: 94°C. for 30 sec.; 55°C. for 2 min.; and 72°C. for 3 min.; followed by post dwell at 2°C.

The amplification initially showed a streak with multiple mRNA. To get the necessary specific primers, the cDNA from the initial amplification was amplified first with only primer (ii) of the 2 Kb coding for the non-repetitive region of the silk protein, which also incorporates the stop codon using 1 μl of the cDNA from the first PCR. This produces single strand cDNA only having a poly d(A) on one end. This new primer only amplifies cDNA coding for silk protein. This produces a selective library for silk proteins. This also gave a streak that amplified preferentially the cDNA from

the silk protein. Next, a PCR method was used whereby 1 μl of the above-described reaction was used with the primer (ii) and poly $d(T)_{20}$. When this was done, there were three distinct mRNA bands formed on an agarose gel with ethidium bromide. These mRNA bands showed that three mRNA's of different sizes formed from the spider silk gene which would code for proteins of about 95 kDa, 190 kDa, and 220 kDa. The 190 kDA and 220 kDA proteins were fortified in natural spider silk, however, all three were formed. The same three proteins are produced both in the clones and the native spider dragline silk as confirmed by electrophoresis. This was important to show cloning of the correct gene. These results convincingly indicate that three start sites existed for this protein as they are homologous for the last 2 Kb according to PCR analysis. The largest of these fragments is about 14 Kb long. The two largest fragments were subsequently cloned. The largest one was cloned by blunt end restriction opening of the pUC18 with Sma I and treated with CIAP as noted above in Example 2. The cDNA was blunt end inserted by ligating this into the vector as shown above in Example 2. This was transformed with supercompetent E. coli XL1 Blue MRF' with kit no. 200230 from Stratagene Cloning Systems, 11011 North Torrey Pines Rd., La Jolla, CA 92037.

Positive transformants were assayed for insertion by checking the size of insertion with a 1% agarose gel. The positive inserts were then tested for the correct insert by using PCR and poly $d(T)_{20}$ primer. The positives were also tested by the antibody methods discussed below. The positives passing the antibody tests for large mRNA were tested using SDS electrophoresis gels and found to give three different proteins also proving multiple start sites. One protein was slightly larger than the 2 Kb piece and the

other two proteins were slightly shorter than native spider silk dragline protein. It was difficult, however, to get these high molecular weight proteins to stain with a Western stain, but this was also true with the native proteins.

Although there was no attempt to put in start codons or insure that the reading frame was correct, these clones produced a large amount of protein. In fact, some produced so much protein that growth was inhibited. To the inventors' knowledge, this was the first time that a culture showed this much synthesis of the target proteins. As further explained in the fermentation section below, it was surprisingly discovered that culture conditions, such as lower temperatures, helped raise protein production. It was discovered that those proteins interfere with isolation of the plasmid DNA for sequencing, thereby making it difficult to get proper sequence while the DNA is coding for protein in the bacteria. However, the last part of the sequence of each of these proteins was the same (except for some minor differences at the amino end where up to 100 bp was deleted from some clones). These clones have the same sequence (the last 1900+ bases) at the carboxy end since they read the same DNA coding region.

Cloning from a Single Site Primer System

As stated earlier, primer (ii) is unique because it codes for the carboxy end of the major ampulate (dragline) silk protein. Nevertheless, it was necessary to develop a method that would get further into the amino direction and hopefully pull out the whole sequence. Two such approaches were developed. One was to use a shotgun method to make DNA clones, which is discussed below. It was believed unlikely that one would be able to clone the whole gene in one

insert and make protein by this method. Because the inventors knew that the carboxy end was unique for other spider silks of interest, they believed a method could be developed for PCR which only had to start with one known unique site. This technique, which is the second approach, involved ligating cassettes to the end of the DNA, although the use of a terminal transferase would have been as effective.

Example 4

To facilitate unique marking at the ends of the DNA whereby PCR primers could be developed that would bind to the site, a number of cassettes from a Takara kit were developed. The cassette systems disclosed below were used.

Cassette 1. Sau3A I Cassette.
5'-GTACATATTGTCGTTAGAACGCGTAATACGACTCACTATAGGGA-3'
3'-CATGTATTACAGCAATCTTGCGCATTATGCTGAGTGATATCCCTCTAG-5'

Cassette 2. EcoR I Cassette.
5'-GTACATATTGTCGTTAGAACGCGTAATACGACTCACTATAGGGAGAG-3'
3'-CATGTATTACAGCAATCTTGCGCATTATGCTGAGTGATATCCCTCTCTTAA-5'

Cassette 3. Hind III Cassette.
5'-GTACATATTGTCGTTAGAACGCGTAATACGACTCACTATAGGGAGA-3'
3'-CATGTATTACAGCAATCTTGCGCATTATGCTGAGTGATATCCCTCTTCGA-5'

Cassette 4. Pst I Cassette.
5'-GTACATATTGTCGTTAGAACGCGTAATACGACTCACTATAGGGAGACTGCA-3'
3'-CATGTATTACAGCAATCTTGCGCATTATGCTGAGTGATATCCCTCTG-5'

Cassette 5. Sal I Cassette.
5'-GTACATATTGTCGTTAGAACGCGTAATACGACTCACTATAGGGAGAG-3'
3'-CATGTATTACAGCAATCTTGCGCATTATGCTGAGTGATATCCCTCTCAGCT-5'

Cassette 6. Xba I Cassette.

5'-GTACATATTGTCGTTAGAACGCGTAATACGACTCACTATAGGGAGAT-3'

3'-CATGTATTACAGCAATCTTGCGCATTATGCTGAGTGATATCCCTCTAGATC-5'

Primer C1.

5'-GTACATATTGTCGTTAGAACGCG-3'

Primer C2.

5'-TAATACGACTCACTATAGGGAGA-3'

Primer (ii).

5'-GGCGAATTCACCCTAGGGCTTGATAAACTGATTGAC-3'

See Isegawa et al., Mol & Cell. Probes 6:467 (1992).

To run this assay, it is necessary to digest the high molecular weight spider genomic DNA at one of the above restriction sites. The restriction digestion procedure is as follows: 2 μl 1 μg/μl genomic DNA; 20 units of an appropriate restriction enzyme (corresponding to one of the six above-mentioned restriction cassettes or others provided the same Restriction Cassette is used with the restriction enzyme); 5 μl 10X buffer for restriction enzyme; distilled water up to a total of 50 μl; and incubate at 37°C. for 3 hr.

This restriction digest is then cleaned and reconcentrated by ethanol precipitation and redissolved in sterile water. The cassette is then ligated to the respective DNA digest. The ligation reaction procedure is as follows: 5 μl genomic DNA digest; 2.5 μl of an appropriate cassette (such as cassettes 1-6 mentioned above) (20 ng/μl); 7.5 μl Takara ligation solution; and incubation for 30 min. at room temperature.

This ligation reaction mix is then cleaned and reconcentrated by ethanol precipitation and redissolved in 5 μl of sterile water. Because the Taq in Takara's kit did not have proofreading activity or high fidelity, reagents and polymerase from the Takara LA PCR kit were used and resulted in very accurate transcription. The protocol used is described below.

The first PCR amplification mix had 2 μl of DNA solution; 1 μl of cassette 7 (primer C1); 1 μl of cassette 9 (primer (ii)); 10 μl of 10X LA Ex Taq polymerase buffer; 1 μl of Ex Taq LA polymerase; 10 μl of dNTPs (2.5 mM each); and water to a total of 100 μl. The PCR conditions were as follows: initial dwell 94°C. for 1 min.; amplification (30 cycles): 94°C. for 30 sec.; 55°C. for 2 min.; and 72°C. for 1-3 min.; and post dwell at 2°C.

After the first PCR amplification, a second PCR was conducted under the same conditions except that the genomic DNA solution was replaced by 1 μl of the first PCR product and cassette 7 (primer C1) was replaced with cassette 8 (primer C2).

An agarose gel of the second PCR product showed bands for three of the cassette systems: Pst I, Hind III and EcoR I. These were faint bands greater than 40 Kb in length and some greater than 100 Kb. While significant streaking of the gel occurred, it was assumed to be due to the extreme length of the PCR products as the inventors were unable to find any reports of PCR of this length. Each of these PCR products was then cut out of the gel and repurified by Gene Clean. These products were blunt end fragments and directly cloned into pUC18 at the Sma I blunt end restriction sites and transformed into E. coli XL1 Blue MRF'. While all of these inserts deleted to some extent when inserted,

they nevertheless produced plasmid clones in excess of 20 Kb (typically about 23 Kb) which was long enough to insert the entire dragline spider silk gene. As discussed below, the E. coli transformants did not grow very well in broth culture because of biochemical problems resulting from high production of silk.

These transformants, like the cDNA transformants previously discussed, did not grow very well and seemed to make cottony masses resembling silk. Because of this, the present inventors set out to determine whether spider silk was being produced. The antibody and hemagglutination tests described below showed the production of large quantities of silk protein. SDS gel electrophoresis detected the presence of three proteins (which from the above-described cDNA work of Example 3 would be expected) and that the largest two fragments were full length matching native spider silk. Western blotting also showed the same results as with the cDNA, i.e., the smaller silk fragment stains very well. Owing to precipitation and other problems, the very large proteins did not Western blot positively like native silks. Example 3 worked like this example. However, in both cases and with native spider silk the larger proteins negatively stained.

There were also some notable problems with the growth of many of these clones -- similar to that observed with cDNA clones but strikingly higher in production. Some of these clones do not grow well in broth cultures like LB broth at 37°C. Interestingly, it was postulated by the inventors that a promoter came with the clones and aided the production rate. One approach to avoid sequencing problems of the full length silk associating with the plasmid DNA (which causes streaking on agarose gels) is to sub-clone into non-protein producing or low molecular weight protein

producing sub-clones. This problem is more severe than that observed with the cDNA clones described above or the multimers described below. These clones like the cDNA clones produce large amounts of protein and can be used for large scale production. The last 2 Kb of the DNA sequence has been already determined to match the 2 Kb insert for which the sequence is completed.

Shotgun Cloning from Genomic Silk DNA

Although this particular method is known for cloning silkworm DNA, the present inventors discovered that this technique is also suitable for spider silk DNA isolation. However, it is expected that up to 50,000 clones will have to be screened by hybridization probes to find a suitable clone which might contain the whole gene. Unlike the other cloning methods discussed above, it is not expected that any of these clones (or only a small number of clones) will produce protein without extensive splicing. Therefore, the present inventors set out to improve this technique. That improvement, which the present inventors developed, involves the use of biotinylated probes, such as those used for cloning, attached to glass beads. It was found that this technique will enrich before cloning sequences having at least a portion of the silk gene. The biotinylated probes select for DNA sequences having the specific region hybridizing to the probe. Therefore, DNA fragment may not have the whole gene. Nevertheless, this technique is used to obtain the spider silk gene and as a starting point for making protein expression clones. Compared to the cloning techniques described in Examples 1-4, these shotgun methods are rudimentary but still suitable methods for cloning Nephila and other spider silk proteins.

Example 5

Using hybridization probes for selecting clones with biotinylated probes is known. For example, a Sigma kit (Cool-1), Sigma Chemical Co., P.O. Box 14508, St. Louis, MO 63178-9916, provides the reagents for final development of dot blots of clones on sigma nitrocellulose membrane blotted according to BioRad procedures. There are many of these procedures and most people skilled in molecular biology have practiced this basic technique.

The technique for concentrating DNA segments hybridizing to biotinylated probes is not known as well. In this system, DNAL (Lake Success, NY) M-280 glass beads were used for pre-enrichment of genomic DNA using the following procedure. First, the biotinylated probes and Dynabeads M-280 Streptavidin were mixed in a microcentrifuge tube. 100 μl of the beads and 100 μl of biotinylated probe (1 μg/μl) were mixed together and allowed to bind for 10 min. at room temperature. The bead is held in a centrifuge tube with a tube magnet and the liquid is gently poured off. The beads are then washed 3 times with TE buffer containing 0.1 M NaCl. 100 μl of genomic DNA that has been pre-denatured at 95°C. for 2 min. is added to the beads. The beads and the DNA are allowed to hybridize for 2 hr. at 42°C. using an equal amount of binding solution that is 2X and consists of 10 mM tris, HCl (pH 7.5), 1 mM EDTA and 2 M NaCl. The temperature is then lowered to room temperature and the beads are washed 3 times in the nybridization solution. The enriched DNA is then eluted by using 0.15 M NaOH containing 0.1 M NaCL. The DNA is concentrated to 5 μl in water and cloned by insertion into the pUC18 vector at the Sma I site. The correct pieces are still selected using various biotinylated probes that bind to spider silk DNA

sequences. Positive clones are sequenced. This technique is very effective but takes quite a bit a work for selection. Enrichment of the DNA can be obtained so that only 500 clones or less need to be screened. Without this enrichment, however, 200,000 to 20 million clones must be screened to obtain a clone having the silk gene.

The Multimerization Process

Using the two primer PCR cloning techniques described in Examples 1 and 2 and more than 20 different primers based upon Nephila type sequences, the 2 Kb inserts were the longest spider silk pieces cloned. Because of this, it was theorized that a different technique would be required to make larger fragments. It was considered necessary that the technique obtain additional sequence information from parts of the protein coding towards the amino end because, with the available information from the protein sequencing, larger fragments were not produced. Although the 2 Kb piece was over 40% of full length, multimerization was considered necessary to increase strength characteristics -- as strength generally varies with the size of the silk polymer. Therefore, the inventors wanted to multimerize the 2 Kb insert to make a larger protein than the natural gene.

The present inventors of the present invention postulated that PCR would make a suitable method to multimerize these inserts as it avoids the repetition of reported sequences. The multimerization processes of the present invention are shown in the following examples.

Construction of PCR fragments with various useful restriction sites was accomplished by modifying the overhangs of the current beginning and ending primers. Other beginning and ending primers like primers (i) and (ii) described above have different restriction sites, in addition to having the stop frame codons deleted so that the protein would continue to be translated into mRNA through the sites, enabling longer constructs to be made. The start codons were left in initially so there would be multiple proteins to help check for deletions and to increase the translation. The primers used to make the differing 2 Kb inserts with unique restriction sites are shown below. These are referred to as primers (xxi) - (xxvi).

Primer (xxi) with a BamH I site.
5'-GGCGGATCCGGATCCATGGCAGCAGCAGCAGCAGCAGCT-3'

Primer (xxii) with a Hind III site.
5'-GGCAAGCTTGGATCCATGGCAGCAGCAGCAGCAGCAGCT-3'

Primer (xxiii) with a Sal I site.
5'-GGCGTCGACGGATCCATGGCAGCAGCAGCAGCAGCAGCT-3'

Primer (xxiv) with BamH I site and no stop codon.
5'-GGCGGSTCCACCCAAGGGCTTGATAAACTGATTGAC-3'

Primer (xxv) with Hind III site and no stop codon.
5'-GGCAAGCTTACCCAAGGGCTTGATAAACTGATTGAC-3'

Primer (xxvi) with Sal I site and no stop codon.
5'-GGCGTCGACACCCAAGGGCTTGATAAACTGATTGAC-3'

Example 6 (a 4 Kb multimer construct)

The 4 Kb construct of this Example was made by PCR of primers (i) and (xxv) with a 2 Kb insert in pUC18. Primers (ii) and (xxii) were used in a separate reaction conducted in accordance with Example 2 above except that a 2 Kb starting plasmid was used instead of genomic DNA. Using the LA (long and accurate) PCR technique, the DNA fragments discovered were the 2 Kb pieces with new restriction sites and two bands representing the entire plasmid.

These bands were subsequently separated by a 1% agarose gel (electrophoresis at 70 V for 90 min. on a 8 cm gel) -- Gene Capsules (Geno Technology, Inc. St. Louis, MO 63108) according to company instructions. The bands were cut with both EcoR I and Hind III restriction enzymes and the vector 2 μg was cut with EcoR I and treated with CIAP as described above in Example 2. Then, one half of each of the two 2 Kb pieces and the vector were repurified by phenol extraction and ethanol precipitations and then dissolved into 10 μl of TE buffer. TE buffer is described in Sambrook et al. A 5 μl aliquot of each was added to a ligase reaction (as described above in the ligation protocol in Example 2) and ligated together. These were electroporated into E. coli XL1 Blue MRF' cells (Kit No. 200230), E. coli TOPP cells (Kit No. 200241) and E. coli Sure cells (Kit No. 200238) using the normal bacterial protocol supplied with the Invitrogen electroporator, Cat. No. S1670-01 obtained from Invitrogen Corp., 3985 B Sorrento Valley Blvd., San Diego, CA 92121. The E. coli cells were obtained from Stratagene Cloning Systems, 11011 North Torrey Pines Road, La Jolla, CA 92037. Sure cells gave better results as fewer transformants had deletions. The successful transformants made proteins of 94 and 188 kDa, the

latter of which is similar to the 190 kDa protein reported in the literature for native spider silk.

Example 7 (a 6 Kb multimer construct)

The 6 Kb construct of this Example was made by PCR of 3 fragments using the procedure described above in Example 2. This procedure consisted of using three 2 Kb constructs and the following primer sets: set 1: primers (i) and (xxiv); set 2: primers (xxi) and (xxv); and set 3: primers (xxii) and (ii).

These were constructed into pUC18 in the same manner as described in the 4 Kb multimer construct of Example 6. While it was discovered that, in some cases with Sure cells, deletion did not occur creating a protein larger than native silk, some deletion did occur in many cases as judged by agarose gels of the PCR product using primers (i) and (ii) and transformant vector DNA. The Sure cells were preferred because they were recombinant deficient. As expected, as the DNA got above a certain size, it became less stable and deleted out repeats.

Example 8 (a 8 Kb multimer construct)

The 8 Kb construct of this Example was made exactly as the 6 Kb construct of the above Example with the exception that 4 separate 2 Kb pieces were made from the following four sets of primers: set 1: primers (i) and (xxiv); set 2: primers (xxi) and (xxv); set 3: primers (xxii) and (xxvi); and set 4: primers (xxiii) and (ii).

These were inserted the same way as the other inserts of Examples 6 and 7. Even though deletions occurred in almost all cases, proteins larger than natural silk were produced indicating that this multimerization

technique could be used to make synthetic silks with superior properties. Using this technique and full length DNA, the sequence could be changed to produce multimer units of natural silk DNA, the final product having much higher molecular weight than normal. Some of these clones produced protein similar in size to or larger than full length natural silk.

Clones from other techniques such as the cDNA and single site systems described above could also be pieced together to make other multimers. Clones up to 800 kDa are possible with the multimerization techniques of this invention using full length clones or pieces therefrom.

Vectors and Production Systems

With the cloning of spider silk proteins, E. coli and pUC18 are the preferred initial production systems. Both have good stable expression of high fidelity and excrete the silk protein through their cell membrane. Although only one example of an expression system is given, the specific inserts coding for natural proteins or multimers derived from them are applicable for use in any vector or genomic incorporation system. Because the potential list of vectors and hosts is prohibitively long, only a few examples are given below.

Bacterial systems

E. coli expression systems are preferred because they have the necessary biochemical machinery to produce very high levels of recombinant proteins and excrete them outside the cell membrane. They are also easy to grow using simple fermentations. Additionally, many of the major problems for protein production with this

system have been overcome as these are among the most common of expression systems. pUC18 is among the most commonly used vectors. Other vectors based upon lytic phage, phagamids, and shuttle vectors are also possible as expression insertion systems in addition to the common man-made plasmids of which pUC is just one. Examples of such plasmids include pBR322, pSP-64, pUR278 and pORF1. Examples of phage vectors include lambda, 12001, lambda gt10, Charon 4a, Charon 40, M13mp19 and other phage modified from natural bacterial phage.

Bacillus expression systems including B. subtilis systems can also be used. These bacteria have the advantage of good secretion by the host, which results in less processing steps and processing costs. Although an expression cassette might be used, it has been found unnecessary with the vector host systems studied thus far. One phagemid that can act as an E. coli and Bacillus shuttle vector is pTZ18R which can be obtained from Pharmacia (Piscataway, NJ).

Many other bacterial systems can be used for expression.

Deposited Clone

A representative clone has been deposited with the American Type Culture Collection (12301 Parklawn Dr. Rockville, Maryland 20852) on June 2, 1995 and given ATCC No. 69832. The deposit consists of E. coli XL1 MRF' cells, strain designation PA21, containing a pUC18 plasmid (23 Kb) with a full length spider silk gene capable of expressing full length Nephila clavipes silk protein.

Yeasts and mold systems

Saccharomyces cerevisiae, Schizosaccharomcyces pombe, Pichia pastoris, Asperillus sp., Hansenula sp., and Streptomyces sp. can be used as expression systems. However, with the exception of Aspergillus and Pichia systems, there is little evidence that these systems will produce more protein than bacteria or be amenable to scale-up. These systems, however, might be more desirable to produce USP or food grade materials since bacterial fermentations have toxins and pyrogens associated with them, whereas many of these yeast and mold systems have already been shown to be safe as food grade materials.

Plant transformation systems

Plant systems can be used for production of transgenic proteins such as silk. Although the quantity of protein may be less than that produced in a microbial system, plant cultivation is rather inexpensive. Agrobacter type transfection systems that allow genetic incorporation into the plant genome can be used. These may be inserted by bacteria such as Agrobacter tumafaciens LB4404 using gene gun insertion, electroporation or a number of other insertion tools. Once inserted, they can be incorporated into the plant genome in a stable and inheritable manner. These plant systems have a number of benefits, such as being conventionally grown and harvested in large tonnages. Farmers have experience raising such industrially important plants as tobacco, soybean, rape seed and other widely grown crops, which are the main plants of interest for silk production. Procedures presently exist for purification of high molecular weight proteins from tobacco and soybeans.

Insect systems

Baculovirus expression systems can be used and are well known for high-level expression of recombinant proteins in insect cell lines. Replication and efficient transfection is accomplished by a number of vectors including pBacPAK6, pBacPAK8 or pBacPAC9. These can be used for high level expression although they may not be as cost effective as other systems.

Other animal systems

There are also many vectors that can be used for insertion into a variety of animals. Although, they are not now vector host systems of commercial value, there might be applications whereby the protein would be helpful in the future.

Fermentation Procedures

The first fermentations of transfected hosts were done in LB broth which consists of 10 grams of bactopeptone, 5 grams of Bacto yeast extract and 5 grams of salt and distilled water to a final concentration of one liter. In this particular broth, either a large amount of precipitate or a cottony mass of spider silk-producing bacteria was observed. This observation was important because it indicated that the proteins were being excreted across the cell membrane. However, these high excretion rates appeared to make the cells somewhat leaky. Therefore, increasing the physiological salt concentration is likely to stabilize the culture.

It was discovered by the inventors that protein production increased at lower temperatures, in particular at room temperature and below. It was also discovered that, at higher temperatures, the protein

disappeared more rapidly (within 5 days) in the fermentation media than at room temperature or below. This phenomenon indicated that a protease was being induced at the higher temperatures around 37°C. This protease activity is noteworthy as many proteases, such as lysozyme and proteinase K, do not seem to degrade spider silk protein. These undesirable metabolic effects are minimized at lower temperatures. This may be due to the induction of shock proteins at lower temperatures.

The composition of the fermentation media was also found to affect the protease activity. For instance, urea-SDS gels of a two day culture did not show protein degradation when grown in LB broth, but when a culture was grown on LB media supplemented with glucose (10 grams of glucose, 10 grams of peptone and 5 grams of yeast extract and distilled water to one liter), there was massive protein degradation after 24 hours. The only difference between the supplemented LB media and the LB broth was that LB broth contained 10 gm/l of NaCl, whereas the supplemented media contained an equivalent amount of glucose.

As a result of the discovery of this protease problem, protease inhibitors were investigated. It was believed that if an inexpensive protease inhibitor could be found and inserted into the culture media, it would be advantageous for fermentation scale-up. The compounds tested included $ZnCl_2$, copper sulfate, disodium EDTA, sodium chloride, boric acid, ethylene glycol bis (B-aminoethyl ether), phenylmethyl sulfonyl fluoride, N,N,N',N'-tetracetic acid, 1,10 phenanthroline, 1,10 phenanthroline iron complex, sucrose, glucose, lactose, fructose, glycerol, peptone and yeast extract.

The most effective inhibitors found were salt additions from NaCl or KCl. Boric acid was also found to be a good inhibitor. None of the other compounds were effective. In fact, the simple sugars, and lactose and glucose in particular, promoted protease activity. Peptone and yeast extract did not affect protease activity. These compounds were tested with AOAC Official Method 969.11, a method for testing proteolytic chillproofing enzymes in beer. To perform this test, 1 ml of the culture was taken and tested. When an active protease was present, the solution cleared in just a few seconds. Protease negative samples showed cloudiness after a ½ hr. at 60°C or overnight at 20°C. This test was used as a quick quality control tool to screen various culture media for its proteolytic enzyme-inducing ability.

Fermentation was attempted using various media. It was found that complex media worked very well. However, acceptable protein production was obtained using 10 times less peptone and yeast extract than contained in LB broth. This simpler and less expensive media produced considerable protein. This media consisted of the following ingredients: 1.2 g dipotassium phosphate, 1.1 g monosodium phosphate, 4.0 g sodium chloride, 0.45 g magnesium sulfate, 2.0 g ammonium sulfate, 0.04 g sodium nitrate, 0.03 g calcium chloride, 0.02 g ferric sulfate, 0.01 g manganese sulfate, 0.01 g boric acid, 0.0005 g sodium molybdate, 0.005 g cobalt chloride, 0.5 g glycine, 1.0 g alanine, 1.0 g yeast extract, 10 g glycerol, distilled water to 1 liter, pH adjusted to 7.0. A wide range of culture media compositions can be used for the fermentations of this invention. These media can range in composition from salts, glycerol (or other carbon sources) and yeast extract or some other source of minor nutrients.

While simpler media is less expensive, it generally results in lower levels of silk protein.

The other main fermentation conditions that must be optimized are oxygen, nutrient level and temperature. Anaerobic conditions at 30°C. has been found to be preferred. In addition, the carbon source should be added at a relatively high level to maximize growth and protein expression. For example, 10 grams of glucose and 10 grams of glycerol per liter has been used.

Antibody Testing

The antibody testing that was developed to determine whether the spider silk protein was expressing in the E. coli host was done with three animal hosts using silkworm silk and spider major ampulate gland silk.

To develop these antibodies, the silkworm protein was taken from fifth star Bombyx mori caterpillars before they spun a cocoon. By selecting such caterpillars, the silk was viscous and gave the caterpillar a translucent appearance that was recognizable. The viscous liquid silk was removed by dissection using aseptic techniques. This silk could then be added to the adjuvant directly. Alternatively, spider silk from the major ampulate gland of the spider could be drawn. However, it was necessary to dissolve the spider silk. This was done by suspending it in 8 M LiBr with heating to 95°C. for 5 min. This spider silk and the silkworm silk were used for making antibody to the silk.

To make the antibody, the LiBr was replaced with 8 M urea and finally in 2M urea by centrifugation. Once the sample is in urea, either sample of silk is mixed (1:10) with 10 ml of Freund's complete adjuvant. This is injected IP into the mice, rabbits or goats to

develop antibodies. On day 21 through 28, the animal was boosted with the silk and Freund's incomplete adjuvant. By the fifth week, it is possible to collect blood weekly and collect the antibody in the serum. The serum was used for running hemagglutination tests or Western blots. Using this procedure for mice, rabbits and goats, blood was taken and the serum separated. This gave polyclonal antibodies to both silkworm and spider silk from each type of animal. These sera were tittered for antibodies and all found to be at least a titer of 256 by standard hemagglutination tests.

Hemagglutination Test

The hemagglutination tests were performed by coating 1% RBCs (Sigma Cat # R-3378). The dissolved silk was added (1mg) to 1 ml of 1% RBCs. This was vortex mixed a few times at room temperature and refrigerated overnight. The next morning, the RBCs are washed by centrifugation in phosphate buffered saline (pH 7.2) three times to remove any non-adhering protein. The sensitized RBCs were then stable in the refrigerator for 2 weeks or longer.

To run a hemagglutination test on the sera, 25 μl of antisera was serially diluted (2 fold dilutions) and 25 μl of sensitized RBCs were added. Control wells were also serially diluted similarly and non-sensitized (25 μl) were added. The microtiter plate was rocked at room temperature for 10 min. and the plates were incubated at room temperature for 90 min. without being disturbed. They were evaluated by the method of Rose and Friedman, (Manual Of Clinical Immunology, 2nd ed., Amer. Soc. Microbiol. (1980)). Many silks have a similar folding structure due to the similarity of their repeating units. Therefore, it was thought that

there might be some cross reactivity due to the tertiary structure being similar. It was found that silkworm antisera cross reacted with spider silk protein sensitized RBC's and that spider silk antisera cross reacted with silkworm protein RBC's. This cross reactivity became a major tool as a culture could be tested against both sets of antibodies with confidence that the silk was not due to another protein that E. coli made.

We also found an additional screening technique that was based upon the coating of RBC's with silk antigens. It was found if we sonicated washed cells, separated the cell membranes and took the supernatant, we could use this instead of the sera by two-fold serial dilutions. Upon putting 25 µl of silk sensitized cells on RBS's they give a classical hemagglutination test that could be used as a first screen of the transformants. It is theorized that the silk protein has sticky ends that will attach to other silk protein in solution and crosslink the RBC's. This would use the same mechanism whereby the silk protein associates and falls out of solution. We could not find any other reference to a protein assay based upon this mechanism. Therefore, we expect that it is a very specific assay for silk and silk-like proteins.

To run the hemagglutination test on the colonies, bacteria cultures (1 ml) were washed 3 times by centrifugation in PBS and brought up into 100 µl of PBS. They were sonicated using a Branson 450 sonicator with a 1/8 inch tip at 40% power and 20% duty cycle for two minutes in an ice bath. This solution is used for sensitizing the RBC's. The assay was run the same way as above except each was for a different bacterial isolate. In all cases, cultures that were successfully producing a silk protein had a titer of at least 16 and

usually 256. This procedure was used to screen the 10 most promising isolates as found by the above agarose gel of the plasmid and blots. In the case of the 2 Kb insert only the best few isolates were saved for further work.

Purification of Silk Proteins

The present invention also encompasses the techniques for purification and spinning the silk. These steps are essential for the processing of the protein into its final form. The protein can be used as a coating, extruded into a fiber, or made into a polymeric film.

The purification of silk protein from the fermentation media can be accomplished by a two step process. First, the bacterial cells and precipitated protein can be removed by continuous centrifugation. The remaining material present in the fermentation broth can be separated by ultrafiltration since most of the protein above a molecular weight of 80,000 is silk. The protein silk streams from the continuous centrifugation and ultrafiltration procedures can then be combined. The bulk of the remaining proteins can be found in the bacterial membranes. By rupturing the bacterial cells using ultrasound, the cells are opened and the silk protein in them is removed.

An important discovery of the present invention is the use of ultrasound to solubilize the spider silk, provided it was not washed and completely dried. This re-solubilized silk protein solution can then be centrifuged to remove the cell membranes. After the cell membranes are removed, the protein can either be further purified by ultrafiltration or spun. In order to spin the silk, it is important to maintain the silk in solution. Prior processes, however, used very harsh

chemicals to maintain silk solubility for spinning operations.

Various compounds will keep the silk protein from re-precipitating prior to the spinning process. These include a variety of salts, lithium salts, sodium and potassium hydroxide, urea phosphate, guanidine hydrochloride, urea, and hexafluoroisopropanol -- all of which dissolve the silk. It was also found by the present inventors that after purification by ultrafiltration, further purification can be effected by alcohol precitation by adding ethanol, methanol, other alcohols or similar solvents. This purified silk protein material could be redissolved by ultrasound or by adding one or more of the above salt compounds. The preferred compounds as determined by cost and environmental considerations for silk protein solubilization are sodium and potassium hydroxide, sodium chloride, potassium chloride and lithium chloride or lithium bromide used in combination with ultrasound or with alcohols for protein purification.

Like other silk proteins, spider silk protein is not easily solubilized. Although there is data that suggests that spider silk may be soluble in harsh chemicals like formic acid (88%), the present inventors found that it caused degradation of full length protein. However, the present inventors found that silk fibers could be resolubilized in LiSCN, LiBr, LiCl, urea, hexafluoroisopropanol, guanidine hydrochloride and similar denaturants. Once the silk proteins are solubilized, less potent denaturants including urea can be used to prevent the protein from re-precipitating. It most likely will be preferred to use soluble protein before irreversibly spinning into a thread. Therefore, silk protein that has been

resolubilized from completely dry silk protein and silk protein that has never been dried completely after being recovered from the fermentation process are recommended for the spinning operations.

Further Processing of Silk Protein into Fabric

General Processing of Silk

Silk protein from silkworms are typically processed in the following manner. To make the silk fibers strong enough for weaving, up to five fibers are twisted together. After the first reeling, the silk is rewound onto skeins, which are twisted together.

The raw silk then goes through several processes called throwing. The skeins are washed and dried and wound on large spools or bobbins. These bobbins are placed on doubling frames where single strands are doubled and twisted together to obtain the desired thread size. This thread is then twisted and drawn out by the spindles of a throwing frame.

Bobbins of silk from the throwing frame are then placed in water and the silk is stretched between rollers. The degree of elongation on the throwing frame affects the fiber diameter. On the stretching frame, the thread is made smooth and even. Before the cloth is finally woven, the thread of thrown silk is boiled to remove any residual water soluble proteins or other gummy substances. Because the boiling step lessens the weight of the silk, the silk is dipped in salts of iron or tin in order to regain some of the lost weight. During this dipping step, the silk fiber takes up some of these salts and becomes heavier but does not lose its luster.

Specific Processing of Spider Silk Protein

The spider silk proteins produced by the above-described methods can be processed into fabrics in the same manner as silkworm proteins. This requires spinning or extruding the protein or protein solutions to obtain silk filaments which may range in diameter from 5 μm to 200 μm or higher. The first step in the process is to concentrate the silk proteins from the fermentation solution. This concentration step can be accomplished by a number of methods including the use of membrane technologies which permit only materials of a given molecular weight range to pass. One disadvantage of using these membranes is cost. Other more cost effective methods to concentrate the silk proteins and remove the host vector include continuous or batch centrifugation. In addition, ultrasound energy can then be used to lyse the bacterial cell wall and allow the silk proteins produced within the cell wall to escape into the aqueous media. To separate the silk proteins from the bacteria cell walls, higher concentrations of salts are favored.

At this point, the protein solution can be precipitated from media by various alcohols. Useful alcohols include methanol, isopropanol and ethanol. The prior art teaches that at this point in the process the silk proteins can be dissolved in lithium salts and organic solvents containing fluorine. However, that procedure is expensive and a severe environmental challenge. In a more preferred embodiment of the present invention, the spider silk proteins are concentrated using alcohols or membrane filters and then maintained in solution in a viscous form by using aqueous solutions of sodium chloride in combination with ultrasound, until they are extruded. If necessary, urea, sodium and potassium hydroxide or lithium salts can be added

as disclosed by prior art processes. However, because of the sodium chloride and ultrasound, only very low concentrations of these materials may need be used. It should be noted that excessively high ultrasound energies, prolonged ultrasound use during purification or high molarity concentrations of lithium salts can reduce the molecular weight of the silk proteins.

Once the protein is extruded using small diameter tubing or other methods that produce small diameter filaments, the protein can be processed in a manner similar to silkworm silk. Once the protein is exposed to air and dried, it is no longer soluble in sodium chloride or by ultrasound.

Similar equipment to that used for silkworm can then be employed to prestretch or throw the silk protein. Boiling, similar to that used for silkworm, can also be used. The majority of the weight loss from boiling is not water soluble protein as with silkworm fibers, but rather the residual salts and water soluble nutrients from the fermentation media. While a 20-25% weight loss is common from raw silkworm silk during this process step, weight loss of less than 5% is expected from the spider silks of the present invention when exposed to the boiling water to clean the final silk or to prepare it for dying.

By using the processes of the present invention, the natural colors of the silk protein can be obtained by selecting primers which encode further into the genomic DNA. White, yellow, pink and light purple colors have been observed with the spider silk proteins produced from the clones and processes of the present invention. The selection for natural color is of value for the manufacture of woven textile fabrics since in many

cases it will eliminate the need and associated cost of color dying.

The spider silk protein filaments can be treated in a manner similar to silkworm silks by winding or twisting two or more threads together to make larger yarns. In addition, these yarns can be interwoven with carbon or graphite fibers, boron or boron coated graphite fibers, or Kevlar® to make woven materials of unusually high strength for body armor and other applications. Conversely, by using smaller yarns, consisting of three to five filaments and only pure spider silk, fabrics with a very smooth feel and luster can be manufactured. The elasticity and other properties of the final weave can in part be controlled by the processing of the filaments or fibers after the extrusion or spinning process. A major variation in process parameters from silkworm processing is the degree of prior elongation or stretch on the individual filaments as they are being drawn or extruded from the initial protein solution or afterwards in a throwing process standard to the silkworm industry.

In addition to the above, the high strength properties of the spider silk protein filaments permit other processing variables. One such process variable is the on line coating of the silk threads using various materials to impart color, increased strength, luster, iridescence and other qualities which increase the marketability of the fabric on the basis of appearance, feel or strength. The on line coating can be accomplished by several methods including running the spider silk filaments through various baths or troughs during the extrusion, rewinding or throwing steps. On line vapor deposition can also be used.

On line vapor deposition of materials onto silk proteins must take into consideration that some residual salts or other fermentation compounds may be present when the filament is initially formed. In addition, these filaments tend to remain wet after being formed unless dried by ovens, fans or other means. In some cases, boiling after extrusion may be preferred to remove all traces of the fermentation media and resolubilization chemicals -- both of which may invoke an allergenic response from the skin when woven into fabrics. Materials that can be vapor deposited onto spider silk proteins include the oxides of tin and titanium. These oxides form a layer on the filaments, the thickness of which depends on the oven conditions. Although titanium coatings may produce higher strength fibers, some people have allergic reactions to titanium dioxide coatings and this may limit its use to applications other than clothing. Tin oxides, however, are GRAS (Generally Recommended As Safe) for human skin contact and therefore can be used in clothing applications.

Films of spider silk protein can be manufactured by several methods including casting wherein the silk protein solution is poured and spread onto sheets or by using rollers. Films may also be modified by the addition of compounds to the protein prior to casting or rolling. This would include the incorporation of active molecules which may act as fragrances, flavors, absorbents or reactants to various biological reagents and weapons. Films may also have colors added during processing or a natural color from a silk clone protein can be selected to impart a natural color.

WHAT IS CLAIMED IS:

1. A process of producing a DNA fragment encoding silk protein, comprising the steps of: selecting target DNA harvested from a silk-producing spider, said target DNA comprising a plurality of repetitive and non-repetitive regions; selecting a single strand DNA primer of at least 10 nucleotides having a DNA sequence that is complementary to a region in said target DNA; and repetitively combining the DNA primer with melted target DNA and incubating said combined DNA primer and target DNA with nucleotides and a DNA polymerase having proofreading ability to produce said DNA fragment, wherein said DNA fragment is complementary to said target DNA and is at least 2 Kb.

2. The process according to Claim 1, comprising the step of using two different DNA primers.

3. The process according to Claims 1 or 2 wherein said target DNA is cDNA made by reverse transcription of full length mRNA coding for spider silk; adding a primer site to the amino end of the first strand cDNA made thereof; and using the poly dT region of the cDNA as a first polymerase priming region.

4. The process according to Claims 1 or 2 wherein a second primer site is created at the unknown end of the DNA using a ligation cassette.

5. The process according to Claims 1 or 2 wherein a second primer site is created at the unknown end of the DNA using a terminal transferase to make a primer site selected from the group consisting of poly dT, poly dA, poly dG and poly dC.

6. The process according to Claims 1 or 2, comprising the step of selecting a spider of the genus Micrathena, Mastophora, Metepeira, Araneus, Argiope, Nephila or Gasteracantha.

7. The process according to Claim 6, comprising the step of selecting a primer DNA represented by sequences (i) - (xx):

(i) GGCGAATTCGGATCCATGGCAGCAGCAGCAGCAGCAGCT;
(ii) GGCGAATTCACCCTAGGGCTTGATAAACTGATTGAC;
(iii) GCATGCACGCATGGTGCATGGATGC;
(iv) TTCGAATTCATGGGCCCTGGACAACAAGGACCATCTGGACCT;
(v) GGAAGGCGGGCAGTGAGCGCAACGCAATTAATG;
(vi) GAYGAYGGNAAYGCNGT;
(vii) TGNTGNCCSGTTCG;
(viii) CGSCGKCGSCCACGSCCSCG;
(ix) GTTAAATGTAAAATCAAGAGTTGCTAA;
(x) GGCCAATCTCTTTTGAGTGCATTTTAA;
(xi) TAAGCAACTCTTGATTTTACATTTAAC;
(xii) TTAAAATGCACTCAAAAGAGATTGGCC;
(xiii) TCAGCAGAATCTGGACAACAAGGCCCA;
(xiv) CCNCGNCCNCTYCC;
(xv) GGTGCAGCAGCAGCAGCTGCWGG;
(xvi) GGTGGTGCCGGACAAGGAGGMTATGGAGGWCTTGGA;
(xvii) GGWGGACGAGGTGGATTA;
(xviii) GATAAAAAGAAATATGCTGCAGAACTTCACTTGGTTCAC;
(xix) CARGCNGGNGCNGCNGSNGGNGGNTTYGGNCC; and
(xx) GGNGGNGGNGCNGGNCARGCNGGNGCNGCNGSNGGNGGNTTYG GNCCNGGNGCNGGNGGN,

wherein N = G, A, T, C; V = G, A, C; B = G, T, C; H = A, T, C; D = G, A, T; K = G, T; S = G, C; W = A, T; M = A, C; Y = C, T; and R = A, G.

8. The process according to Claim 7, wherein said target DNA is selected by hybridization to a DNA probe

having the sequences (i) - (xx) which is reversibly bound to a support to enrich for the silk-encoding DNA fragments.

9. The process according to Claim 6, wherein said DNA fragment is at least 5 Kb.

10. A DNA sequence encoding spider silk protein, said DNA sequence comprising a plurality of repetitive and non-repetitive regions and having a length of at least 2 Kb.

11. The DNA according to Claim 10, wherein said DNA sequence has a length of at least 5 Kb.

12. The DNA according to Claims 10 or 11, wherein said spider is of the genus Micrathena, Mastophora, Metepeira, Araneus, Argiope, Nephila or Gasteracantha.

13. The DNA according to Claim 11, wherein said spider is Nephila clavipes.

14. The DNA according to Claim 12, wherein said DNA comprises the sequence illustrated in Figure 1.

15. A multimerization process, comprising the steps of: selecting a target DNA encoding silk protein harvested from a silk-producing spider, said target DNA comprising a plurality of repetitive and non-repetitive regions; selecting a first pair of different DNA primers, said first pair of DNA primers both being complementary to a region in said target DNA, at least one of said first pair of DNA primers being represented by the sequences (i) - (xxvi):

(i) GGCGAATTCGGATCCATGGCAGCAGCAGCAGCAGCAGCT;

(ii) GGCGAATTCACCCTAGGGCTTGATAAACTGATTGAC;

(iii) GCATGCACGCATGGTGCATGGATGC;

(iv) TTCGAATTCATGGGCCCTGGACAACAAGGACCATCTGGACCT;

(v) GGAAGGCGGGCAGTGAGCGCAACGCAATTAATG;

(vi) GAYGAYGGNAAYGCNGT;

(vii) TGNTGNCCSGTTCG;

(viii) CGSCGKCGSCCACGSCCSCG;

(ix) GTTAAATGTAAAATCAAGAGTTGCTAA;

(x) GGCCAATCTCTTTTGAGTGCATTTTAA;

(xi) TAAGCAACTCTTGATTTTACATTTAAC;

(xii) TTAAAATGCACTCAAAAGAGATTGGCC;

(xiii) TCAGCAGAATCTGGACAACAAGGCCCA;

(xiv) CCNCGNCCNCTYCC;

(xv) GGTGCAGCAGCAGCAGCTGCWGG;

(xvi) GGTGGTGCCGGACAAGGAGGMTATGGAGGWCTTGGA;

(xvii) GGWGGACGAGGTGGATTA;

(xviii) GATAAAAAGAAATATGCTGCAGAACTTCACTTGGTTCAC;

(xix) CARGCNGGNGCNGCNGSNGGNGGNTTYGGNCC; and

(xx) GGNGGNGGNGCNGGNCARGCNGGNGCNGCNGSNGGNGGNTTYG GNCCNGGNGCNGGNGGN;

(xxi) GGCGGATCCGGATCCATGGCAGCAGCAGCAGCAGCAGCT;

(xxii) GGCAAGCTTGGATCCATGGCAGCAGCAGCAGCAGCAGCT;

(xxiii) GGCGTCGACGGATCCATGGCAGCAGCAGCAGCAGCAGCT;

(xxiv) GGCGGSTCCACCCAAGGGCTTGATAAACTGATTGAC;

(xxv) GGCAAGCTTACCCAAGGGCTTGATAAACTGATTGAC; and

(xxvi) GGCGTCGACACCCAAGGGCTTGATAAACTGATTGAC

producing a first DNA fragment by repetitively combining said first pair of DNA primers with melted target DNA and incubating said combined DNA primers and target DNA with nucleotides and a DNA polymerase having proofreading ability to produce said first DNA fragment, said first DNA fragment being complementary to said target DNA and at least 2 Kb; said multimerization process further comprising selecting a

second pair of different DNA primers, at least one of said second pair of DNA primers being different than both of the sequences of said first pair of DNA primers, and at least one of said second pair of DNA primers being represented by the sequences (i) - (xx); producing a second DNA fragment by repetitively combining said second pair of DNA primers with melted target DNA and incubating said combined DNA primers and target DNA with nucleotides and a DNA polymerase having proofreading ability to produce said second DNA fragment, said second DNA fragment being different than said first DNA fragment and also being complementary to said target DNA, said second DNA fragment being at least 2 Kb; restricting said first and second DNA fragments; and recombining the restricted portions of said first and second DNA fragments into a multimerized DNA, said multimerized DNA encoding spider silk protein and being at least 4 Kb.

16. The multimerization process according to Claim 15, wherein all DNA primers are represented by sequences (i) - (xxvi).

17. The multimerization process according to Claim 16, wherein all DNA primers are different.

18. The multimerization process according to any of Claims 15-17, wherein said multimerized DNA is at least 6 Kb.

19. The multimerization process according to Claim 18, wherein said multimerized DNA is at least 8 Kb.

20. A process of producing silk protein, comprising the steps of: selecting a DNA according to Claim 12, inserting said DNA into an expression vector; transfecting host cells with said expression vector;

fermenting said transfected host in culture media to produce silk protein; and recovering said silk protein.

21. The process according to Claim 20, wherein said culture media contains protease inhibitor.

22. The process of producing silk protein according to Claim 21, further comprising the steps of: applying ultrasound energy to rupture the host cells; applying ultrasound energy to resuspend silk protein; and centrifuging said ruptured host cells to separate cell membranes from said silk protein.

23. The process of producing silk protein according to Claim 22, further comprising the steps of purifying the silk protein by ultrafiltration or alcohol precipitation.

24. The process for spinning silk protein comprising the steps of concentrating silk protein purified according to Claim 23; drawing a fiber of concentrated silk protein; spinning silk fibers to produce a silk thread; and washing the silk thread to remove any solubilization reagents.

25. The process for spinning silk according to Claim 24, wherein the solubilization reagents are selected from the group consisting of sodium hydroxide, potassium hydroxide, hexafluoroisopropanol, guanidine hydrochloride, urea, urea phosphate, lithium salts, organic solvents, ammonium sulfate, acetic acid, phosphoric acid, dichloroacetic acid, formic acid and sulfuric acid.

26. The process for spinning silk according to Claim 24, further comprising the step of coating said silk fibers or threads with oxides of tin or titanium.

27. A fabric comprising silk threads made according to Claim 24.

28. A fabric according to Claim 27, further comprising silkworm, Kevlar®, graphite or carbon fibers.

```
                9                  18                  27                  36                  45                  54
5' ACA GGA AAC AGC TAT GAC CAT GAT TAC GAA TTC GGA TCC ATG GCA GCA GCA GCA
                                                       Met Ala Ala Ala Ala

               63                  72                  81                  90                  99                 108
GCA GCA GCT GGA GGT GCC GGA CAA GGA GGA TAT GGA GGT CTT GGA AGC CAG GGT
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly

              117                 126                 135                 144                 153                 162
GCT GGA CGA GGT GGA CAA GGT GCA GGC GCA GCC GCA GCA GCA GCC GGA GGT GCT
Ala Gly Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala

              171                 180                 189                 198                 207                 216
GGA CAA GGA GGA TAC GGA GGT CTT GGA AGC CAA GGT GCT GGA CGA GGA GGA TTA
Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Leu

              225                 234                 243                 252                 261                 270
GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA GCA GCA GCT GGA GGT GCC GGA CAA
Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln

              279                 288                 297                 306                 315                 324
GGA GGA CTA GGT GGA CAA GGT GCT GGA CAA GGA GCT GGA GCA GCC GCT GCA GCA
Gly Gly Leu Gly Gly Gln Gly Ala Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala

              333                 342                 351                 360                 369                 378
GCT GGT GGT GCC GGA CAA GGA GGA TAT GGA GGT CTT GGA AGC CAA GGT GCT GGA
Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly

              387                 396                 405                 414                 423                 432
CGA GGT GGA CAA GGT GCA GGC GCA GCC GCA GCA GCA GCC GGA GGT GCT GGA CAA
Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln

              441                 450                 459                 468                 477                 486
GGA GGA TAC GGT GGA CAA GGT GCC GGA CAA GGA GGC TAT GGA GGA CTT GGA AGT
Gly Gly Tyr Gly Gly Gln Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser

              495                 504                 513                 522                 531                 540
CAA GGT GCT GGA CGA GGA GGA TTA GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA
Gln Gly Ala Gly Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala

              549                 558                 567                 576                 585                 594
GCA GCA GCT GGA GGT GCC GGA CAG GGA GGA TTA GGT GGA CAA GGT GCT GGA CAA
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Leu Gly Gly Gln Gly Ala Gly Gln
```

FIGURE 1, PAGE 1
DNA SEQUENCE ENCODING SPIDER SILK PROTEIN

```
          603         612         621         630         639         648
GGA GCT GGA GCA GCC GCT GCA GCA GCT GGT GGT GCC GGA CAA GGA GGA TAT GGA
Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly

          657         666         675         684         693         702
GGT CTC GGA AGC CAA GGT GCA GGA CGA GGT GGA TCA GGT GGA CAA GGG GCA GGT
Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Ser Gly Gly Gln Gly Ala Gly

          711         720         729         738         747         756
GCA GCA GCA GCA GCA GCT GGA GGT GCC GGA CAA GGA GGA TAT GGA GGT CTT GGA
Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly

          765         774         783         792         801         810
AGC CAA GGT GCA GGA CGA GGT GGA TTA GGT GGA CAA GGT GCA GGT GCA GCA GCA
Ser Gln Gly Ala Gly Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala

          819         828         837         846         855         864
GCA GCA GCA GCT GGA GGT GCT GGA CAA GGA GGA TAC GGT GGT CTT GGT GGA CAA
Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gly Gln

          873         882         891         900         909         918
GGT GCC GGA CAA GGT GGC TAT GGA GGA CTT GGA AGC CAA GGT GCC GGA CGA GGA
Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly

          927         936         945         954         963         972
GGA TTA GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA GCA GCA GCT GGA GGT GCC
Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Ala Gly Gly Ala

          981         990         999         1008        1017        1026
GGA CAA GGA GGA CTA GGT GGA CAA GGT GCT GGA CAA GGA GCT GGA GCA GCC GCT
Gly Gln Gly Gly Leu Gly Gly Gln Gly Ala Gly Gln Gly Ala Gly Ala Ala Ala

          1035        1044        1053        1062        1071        1080
GCA GCA GCT GGT GGT GCC GGA CAA GGA GGA TAT GGA GGT CTT GGA AAC CAA GGT
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Asn Gln Gly

          1089        1098        1107        1116        1125        1134
GCT GGA CGA GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA GCA GCT GGA GGT GCT
Ala Gly Arg Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala

          1143        1152        1161        1170        1179        1188
GGA CAA GGA GGA TAT GGA GGT CTT GGA AGC CAA GGT GCA GGA CGA GGT GGA TTA
Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Leu
```

FIGURE 1, PAGE 2
DNA SEQUENCE ENCODING SPIDER SILK PROTEIN

```
          1197       1206       1215       1224       1233       1242
GGT GGA CAA GGT GCA GGT GCA GCA GCA GCA GCA GCT GGA GGT GCT GGA CAA GGA
Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly

          1251       1260       1269       1278       1287       1296
GGA TAC GGT GGT CTT GGT GGA CAA GGT GCC GGA CAA GGA GGC TAT GGA GGA CTT
Gly Tyr Gly Gly Leu Gly Gly Gln Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu

          1305       1314       1323       1332       1341       1350
GGA AGC CAA GGT TCT GGT CGA GGA GGA TTA GGT GGA CAA GGT GCA GGT GCA GCA
Gly Ser Gln Gly Ser Gly Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala

          1359       1368       1377       1386       1395       1404
GCA GCA GCA GCT GGA GGT GCT GGA CAA GGA GGA TTA GGT GGA CAA GGT GCT GGA
Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Leu Gly Gly Gln Gly Ala Gly

          1413       1422       1431       1440       1449       1458
CAA GGA GCT GGA GCA GCC GCT GCA GCA GCT GGT GGT GCT GGA CAA GGA GGA TAT
Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr

          1467       1476       1485       1494       1503       1512
GGA GGT CTT GGA AGC CAA GGT GCT GGA CGA GGT GGA CAA GGT GCA GGC GCA GCC
Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Gln Gly Ala Gly Ala Ala

          1521       1530       1539       1548       1557       1566
GCA GCA GCA GCC GGA GGT GCT GGA CAA GGA GGA TAT GGT GGT CTT GGT GGA CAA
Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gly Gln

          1575       1584       1593       1602       1611       1620
GGT GTT GGC CGA GGT GGA TTA GGT GGA CAA GGT GCA GGC GCA GCG GCA GCT GGT
Gly Val Gly Arg Gly Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Gly

          1629       1638       1647       1656       1665       1674
GGT GCT GGA CAA GGA GGA TAT GGT GGT GTT GGT TCT GGG GCG TCT GCT GCC TCT
Gly Ala Gly Gln Gly Gly Tyr Gly Gly Val Gly Ser Gly Ala Ser Ala Ala Ser

          1683       1692       1701       1710       1719       1728
GCA GCT GCA TCC CGT TTG TCT TCT CCT CAA GCT AGT TCA AGA GTT TCA TCA GCT
Ala Ala Ala Ser Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala

          1737       1746       1755       1764       1773       1782
GTT TCC AAC TTG GTT GCA AGT GGT CCT ACT AAT TCT GCG GCC TTG TCA AGT ACA
Val Ser Asn Leu Val Ala Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Ser Thr
```

FIGURE 1, PAGE 3
DNA SEQUENCE ENCODING SPIDER SILK PROTEIN

```
          1791        1800        1809        1818        1827        1836
ATC AGT AAC GTG GTT TCA CAA ATA GGC GCC AGC AAT CCT GGT CTT TCT GGA TGT
Ile  Ser  Asn  Val  Val  Ser  Gln  Ile  Gly  Ala  Ser  Asn  Pro  Gly  Leu  Ser  Gly  Cys

          1845        1854        1863        1872        1881        1890
GAT GTC CTC ATT CAA GCT CTT CTC GAG GTT GTT TCT GCT CTT ATC CAG ATC TTA
Asp  Val  Leu  Ile  Gln  Ala  Leu  Leu  Glu  Val  Val  Ser  Ala  Leu  Ile  Gln  Ile  Leu

          1899        1908        1917        1926        1935        1944
GGT TCT TCC AGC ATC GGC CAA GTT AAC TAT GGT TCC GCT GGA CAA GCC ACT CAG
Gly  Ser  Ser  Ser  Ile  Gly  Gln  Val  Asn  Tyr  Gly  Ser  Ala  Gly  Gln  Ala  Thr  Gln

          1953        1962        1971        1980        1989        1998
ATC GTT GGT CAA TCA GTT TAT CAA GCC CTA GGG TGA ATT CGA GCT CGG TAC CCG
Ile  Val  Gly  Gln  Ser  Val  Tyr  Gln  Ala  Leu  Gly  ***

     2004
GGG ATC 3'
```

FIGURE 1, PAGE 4
DNA SEQUENCE ENCODING SPIDER SILK PROTEIN

INTERNATIONAL SEARCH REPORT

Inter[illegible]plication No
PCT/US 96/13767

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12Q1/68 D01F4/02 C07K1/34 C07K14/435

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N C12Q D01F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP,A,0 452 925 (THE UNIVERSITY OF WYOMING) 23 October 1991 cited in the application	10,12, 13,20
A	see page 3, line 20 - page 4, line 25 see page 5, line 1 - page 12, line 10; examples 3-6	21-28

X	WO,A,91 16351 (THE UNITED STATES OF AMERICA, SECRETARY OF THE ARMY, THE PENTAGON) 31 October 1991 cited in the application	10,12, 13,20
A	see page 4, line 11 - page 7, line 8 see page 8, line 17 - page 11, line 10 see page 14, line 31 - page 18, line 17 see page 22, line 11 - page 24, line 27; examples 5,6	1-9

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 January 1997	14.02.97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	**Authorized officer** Montero Lopez, B

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 87, no. 18, September 1990, WASHINGTON US, pages 7120-7124, XP002023818 MING XU ET AL.: "Structure of a protein superfiber: Spider dragline silk" cited in the application see page 7120, right-hand column, paragraph 1 see page 7121, right-hand column, paragraph 3 - page 7123, left-hand column, paragraph 3; figure 2 see page 7123, right-hand column, last paragraph	10,12, 13,20
X	JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 267, no. 27, 25 September 1992, MD US, pages 19320-19324, XP002023819 MICHAEL B. HINMAN ET AL.: "Isolation of a clone encoding a second dragline silk fibroin" cited in the application see page 19320, right-hand column, paragraph 2 - paragraph 3; figure 1 see page 19321, left-hand column, paragraph 1 - page 19323, left-hand column, paragraph 1	10,12, 13,20
A	JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 269, no. 9, 4 March 1994, MD US, pages 6661-6663, XP002023820 RICHARD BECKWITT ET AL.: "Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae)" cited in the application see the whole document	1-19
A	ANNUAL MEETING OF THE AMERICAN SOCIETY OF ZOOLOGISTS, AMERICAN MICROSCOPICAL SOCIETY, ANIMAL BEHAVIOR SOCIETY, THE CRUSTACEAN SOCIETY AND THE INTERNATIONAL ASSOCIATION OF ASTACOLOGY, ATLANTA, GEORGIA, USA, DECEMBER 27-30, 1991. AM ZOOL 31 (5). 1991. , XP000614782 BECKWITT R: "AMPLIFICATION AND ANALYSIS OF SPIDER SILK GENES USING POLYMERASE CHAIN REACTION." see abstract no. 168	1-14

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 96/13767

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-452925	23-10-91	JP-A-	6098771	12-04-94
WO-A-9116351	31-10-91	AU-A-	7691191	11-11-91
		US-A-	5245012	14-09-93